DOCUMENT RESUME

ED 337 468 TM 017 281

AUTHOR Shepard, Lorrie A.; Dougherty, Katharine Cutts
TITLE Effects of High-Stakes Testing on Instruction.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.; Spencer Foundation, Chicago, Ill.
PUB DATE Apr 91
NOTE 39p.; Paper presented at the Annual Meetings of the American Educational Research Association (Chicago, IL, April 3-7, 1991) and the National Council on Measurement in Education (Chicago, IL, April 4-6, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Attitude Measures; Educational Practices; Elementary Education; *Elementary School Teachers; Grade 3; Grade 5; Grade 6; *Instructional Effectiveness; Questionnaires; *Standardized Tests; *Teacher Attitudes; Teaching Methods; Test Coaching; Testing Problems; Test Results; *Test Use
IDENTIFIERS *Measurement Driven Instruction; *Testing Effects

ABSTRACT

The effects of standardized testing on instruction were studied in two school districts with high-stakes testing. The present study was part of a larger research project concerned with the effect of testing on instruction and student learning. A total of 360 teachers in grades 3, 5, and 6 in approximately 100 schools in two districts answered a questionnaire addressing test preparation/coaching practices and the effects of testing on instruction. A limitation of the study was the overall response rate of 42%, suggesting that the respondents were not necessarily representative of all teachers. Teachers reported that they felt pressured to improve test scores by the district administration and the media. Because of the importance of testing, teachers gave greater emphasis to basic skills instruction. They felt that content not tested suffered because of the focus on the standardized tests. Testing further distorted teaching because of the extensive time given to test preparation. While it was agreed that flagrant cheating was rare, practices that would clearly boost test scores, such as rephrasing questions, were considered to occur more frequently. Teachers were aware of extensive use of test results for external purposes such as comparisons of schools or districts. In open-ended questions, teachers could point to many benefits from standardized testing, but they felt that these were outweighed by the drawbacks. Fourteen tables summarize the questionnaire results. A 13-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LORRIE SHEPARD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Effects of High-Stakes Testing on Instruction

Lorrie A. Shepard
Katharine Cutts Dougherty
University of Colorado at Boulder

ED337468

Traditionally standardized tests of achievement were used to report to parents and to monitor state and district trends. Such tests had almost no effect on instruction because teachers paid so little attention to them. According to a national survey conducted by Goslin in 1967, teachers only infrequently used the results of standardized tests and reported virtually no influence of test content on teaching methods or course content. This picture changed, however, beginning with the minimum competency testing movement in the 1970s and educational reforms in the 1980s. It was the explicit intention of reformers in recent decades to change instruction by imposing tests.

The debate about the positive or negative effects of testing on instruction is now a familiar one. Proponents of test-based reforms saw tests as a means for external agencies, such as the state or district, to set standards and assure their attainment. This position was best exemplified by Popham's (1987) advocacy of measurement-driven instruction. Popham argued that if tests measure important skills and have sufficiently high stakes, they will serve as "instructional magnets" thus dramatically improving the efficiency and effectiveness of instruction. Opponents of high-stakes, externally-mandated tests protested, however, that such tests would have a deleterious effect on the quality of education by narrowing the focus of instruction to only tested content and by encouraging presentation of content in fragmented bits (Bracey, 1987).

Evidence to support the positive claims for measurement-driven instruction comes primarily from high-stakes tests themselves. For example, Popham, Cruse, Rankin, Sandifer, and Williams (1985) and Popham (1987) pointed to the steeply rising passing rates on minimum competency tests as demonstrations that MDI had improved student learning. In South Carolina, for instance, the percentage of first graders passing the state's basic skills assessment in reading increased from 70% in 1981 to 80% in 1984. In first grade mathematics the passing rate increased from 68% to 81% (Popham et al., 1985). Similar gains were reported in other grades levels and in other states. Popham et al. (1985) also observed instructional changes being made in response to high-stakes tests, such as the use of test specifications as instructional guides, and noted that the greatest gains in achievement occurred where there had been the greatest efforts to direct instruction toward the test.

It is possible, however, for test scores to go up without there being a commensurate gain in learning. The possibility that high-stakes testing programs can produce inflated or spurious results was suggested, for example, by the Cannell's (1987) finding that all 50 states claimed to be above average. Independent evidence of achievement trends from the National Assessment provides both good news and

---

Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April 1991.
The research reported here was supported by grants from the Spencer Foundation and from the Office of Educational Research and Improvement, Department of Education (OERI/ED)

TM017281

bad news for proponents of measurement-driven instruction. Student performance has improved, according to the NAEP, in basic reading and math skills during the period of high-stakes testing. At the same time, however, there has been no gain or a decrease in higher-order, advanced skills. Many attribute this pattern of results to the negative influence of standardized testing on teaching and learning.

Evidence documenting the negative influence of testing on instruction is limited to a few studies. Darling-Hammond and Wise (1985) reported that teachers in their study were pressured to "teach the test." Teachers gave specific examples, such as being asked to write materials similar to the test to give their students practice, and giving up on essay tests so that classroom experiences would more closely parallel standardized tests. State and district testing directors interviewed by Shepard (1990a, 1990b) acknowledged that teachers generally "spend more time teaching the specific objectives on the test(s) than they would if the tests were not required." Testing directors were divided, however, as to whether they saw the influence of tests as positive or negative. Some believed that focusing instruction was beneficial because it ensured that essential skills were taught. Others regretted that higher-order thinking skills and subjects like science and social studies suffered because of the emphasis on basics.

Smith, Edelsky, Draper, Rottenberg, and Cherland (1990) conducted an 18-month observational study of testing effects in two schools. Their key findings, summarized in Rottenberg and Smith (1990), were as follows. (1) "External testing reduces the time available for ordinary instruction." (p. 5) Smith et al. estimated that in elementary grades test-related activities including test preparation, internal testing, and recovery from testing as well as the test administration per se take over 100 hours, equivalent to a full three to four weeks of school. (2) "In high stakes environments, schools neglect material that the external tests do not include." (p. 6) Specifically, teachers spend little time on science, social studies, and writing, concentrating instead on reading, word recognition, recognition of errors in spelling, usage, punctuation, and arithmetic operations. (3) "External testing encourages use of instructional methods that resemble testing." (p. 8) For example in one school where test scores fell just short of a year's growth in language, the principal created a daily review program that required pupils to answer multiple-choice questions on grammar, usage, punctuation, and capitalization.

Romberg, Zarinnia, and Williams (1989) reported survey results for a national sample of eighth-grade mathematics teachers. Teachers were asked about the uses made of state and district test results and about the influence of tests on teaching. Teachers said that as a result of testing they had increased instruction in areas such as basic skills (30% of the 354 respondents), pencil and paper computation (25%), topics emphasized on the test (24%), problem solving (23%), and direct instruction to the whole class (16%). Instructional activities that decreased in response to testing were extended project work (19%), use of calculators (16%), topics not emphasized on the test (14%), use of computers (13%), and cooperative learning (10%). Romberg et al. concluded that instructional changes promoted by standardized testing were antithetical to the kinds of instructional changes sought by the mathematics community and represented in the NCTM Curriculum and Evaluation Standards for School Mathematics (1989).

The present study was part of a larger research project concerned with both the effect of testing on instruction and on student learning. In the larger study (Koretz, Linn, Dunbar, & Shepard, 1991) the effect of standardized testing on student achievement was examined by administering a variety of other measures,

both standardized tests and alternative tests, to see if students really knew what they appeared to know based on their publicly-reported high-stakes test scores. The study reported here addressed the effects of standardized testing on instruction. It was intended in particular to extend the work of Smith et al. (1990) and Romberg et al. (1989). What do elementary teachers in two high-stakes districts perceive to be the influences of testing on their teaching? Are high-stakes uses or pressure to raise test scores associated with instructional changes? What specific kinds of instructional activities do they change in response to tests? We were also interested in the kinds of uses made of test results, and in teachers' perceptions of the importance given to test results? Finally, how much time is spent testing and in preparing for tests? What kinds of test preparation strategies do teachers use? Are teachers aware of cheating or questionable testing practices?

METHODS

Teachers in two high-stakes school districts were surveyed by means of a questionnaire addressing test-preparation practices and the effects of testing on instruction. Sampling procedures, response rates, and the survey instrument are described in the following sections.

Sampling and response rates

The high-stakes districts in this study were the same ones described in Koretz, Linn, Dunbar, and Shepard (1991). They were selected according to the following criteria: (1) their standardized testing programs were described by their Directors of Testing and Research as having very high stakes; (2) they represented different geographic regions of the country; (3) they administered very popular but different standardized tests; (4) they had sufficiently large minority populations; and (5) the Directors of Testing and Research and district personnel were willing to participate. For purposes of the Koretz et al. study it was also essential that only large districts be selected to allow for different random samples of classrooms to take different alternative tests. District A is located in the southeast and has an enrollment greater then 60,000 with approximately 30 percent minority students. District B is located in the southwest and has an enrollment greater than 50,000. About three-quarters of District B's schools have minority enrollments above 70 percent.

Within the two high-stakes districts the intended sample of teachers consisted of two groups. First, we wished to include all the third-, fifth-, and sixth-grade teachers in the random samples of schools selected for the pupil testing part of the project. In addition, because we feared that teachers' responses might be biased by the effect of giving an additional test as part of the research study, we identified a second random sample of schools and targeted the third-, fifth-, and sixth-grade teachers in those schools as well.

Because neither district had directories of teachers names available, questionnaires were distributed through school principals. Individual stamped envelopes were attached to each questionnaire to enable returns directly to the researchers. Principals were also asked to return a form with exact counts of the third-, fifth-, and sixth-grade teachers in each school.

Because of the constraints of the student testing portion of the research, the distribution of the teacher questionnaires was poorly timed. To avoid illegitimate practice on parallel forms of the districts' regular standardized tests, the parallel

forms test could not be administered until after the districts' normal standardized testing period in April. It was also decided that the teacher questionnaire should not be administered until after all standardized testing had been completed for the year. As a result the teacher questionnaires were distributed the third week in May, 1990 with only two weeks remaining until the end of the school year. This timing undoubtedly reduced return rates.

To calculate response rates we estimated the numbers of teachers in the intended samples from the counts sent to us by responsive principals plus the estimated numbers of teachers in schools where the principal did not respond but at least one teacher did. We did not include in the estimate counts for those schools where we never heard from the principal or the teachers, because we believed in these cases that the questionnaires were never distributed to teachers. In District A, 80 teachers responded for a response rate of 37%. In District B, the response rate was helped by a follow-up letter sent to principals in August just as teachers were returning to school; 280 teachers responded from District B, or 44% of the intended sample. Altogether 360 teachers responded, representing approximately 100 schools, for a combined response rate of 42%.

Teacher questionnaire

A four-page questionnaire was developed to include questions in the following categories: Pressure to improve test scores, instructional effects, preparation for tests, controversial testing practices, uses of test data, positive and negative effects of standardized testing, and background information on teachers and schools.* Specific questionnaire items are displayed in the results section of the paper. We also provided teachers with two open-ended questions asking them to report on specific examples of positive or negative influences of standardized tests on their teaching or on students in their classroom.

RESULTS

Factor analysis

To simplify reporting of results we wished to consider groupings of similar items. Factor analysis was used to check on the validity or meaningfulness of intended subscales. Principle axis factoring with iterated communalities was applied; because the factors were expected to be correlated an oblique rotation was specified. Using a criterion of an eigenvalue greater than one, it was possible to extract 20 factors. However, examination of the scree plot suggested either 11 or nine factors. We eventually settled on a nine-factor solution because this produced more interpretable results without items in the last factors loading on multiple factors. The results of the factor analysis are reported in Table 1.

---

*In devising questions on instructional changes we drew specific examples from the study by Romberg et al. of secondary mathematics and developed our own exemplars for reading and for elementary grades mathematics. We are also grateful to Mary Lee Smith for suggesting test preparation and instructional effects items based on the Smith et al. study. We thank Evelyn Brzezinski and Michael Hiscox of Interwest Applied Research for their advice and help in designing the final format of the questionnaire.

Whereas in advance we had anticipated that there might be one general factor explaining attitudes toward tests, pressure for accountability , and instructional effects, instead we found that there were discrete factors corresponding to each of the subparts of the questionnaire. In addition three of the questionnaire subparts were each further subdivided into two distinct but correlated factors. Factor 1 included all of the items about controversial testing practices; these items were used then to create the Controversial Testing Practices subscale (Controv). The specific items comprising this scale are listed in Table 6. Factor 2 and Factor 7 were composed of subsets of items from the Testing and Instruction part of the instrument (see Tables 3 & 4). We had expected that when teachers emphasized basic skills instruction, vocabulary lists and the like, they would necessarily give less attention to higher-order skills and activities such as extended project work in mathematics. Instead, items in Factor 2 which we labeled Skills Instruction (Skillins) and those in Factor 7 which we called Divergent Instructional Practices (Divergen) formed two distinct factors. Interestingly these two factors were slightly positively correlated (r = .2) rather than inversely related.

Factor 3 corresponded exactly to the Test Preparation (Testprep) subscale (see Table 5). Factors 4 and 9 were subdivisions of the Test Use questions (see Tables 7 & 8). Because one set of items referred to uses of tests such as newspaper rankings, comparing districts and comparing schools we called it the External Uses of Tests scale (Extuse). The other set, consisting of items such as evaluating teachers, evaluating principals, and allocating resources, was labelled Internal Uses of Tests (Intuse).

The positive and negative statements devised to assess the effects of standardized tests were sorted precisely into two separate factors, Factors 5 and 6. The Pro Standardized Testing scale (Pro-tests) included items such as, "Standardized testing is helping schools improve," and "Tests give me important feedback about how well I am teaching in each curricular area." The scale we identified as Measurement-Driven Instruction (Measdriv) was marked by items such as, "There are exciting new curriculum developments (e.g., whole language, minds-on-science) that I haven't implemented because they aren't compatible with the tests we're measured by," and "I tend to drill students on basic skills because that is the only way I can be sure they will get what they really need to know."

Five items about sources of pressure to raise test scores comprised Factor 8 (Pressure). Only four items in the entire questionnaire did not fit with their intended subscale: writing essays, use of multiple choice exercises, timed computational exercises, and the use of test scores to promote or retain students. These items were omitted from the analysis of subscales and are reported on separately.

Pressure to raise test scores

Consistent with the selection of these districts as high-stakes sites, teachers reported that they are under substantial, even great pressure to raise test scores. As shown in Table 2, 53% said that they feel great pressure from the district administration or board of education to raise scores. Only 8% responded that the pressure from the district was slight or non-existent. The next strongest source of pressure was the newspaper or media. Most teachers did not, however, report feeling pressure from parents or from other teachers. The results were generally consistent across the two districts except that teachers in District B reported significantly greater pressure from other teachers.

Even within these districts where testing was thought to be pervasively high-stakes, variation in perceived pressure was associated with other effects. For example, as shown later in Table 13, the Pressure scale was correlated with teachers' reports of measurement-driven instructional practices ($r = .30$), with internal (to the district) uses of test results ($r = .29$), and with the extent of test preparation practices ($r = .24$).

Instructional activities emphasized because of testing

Data in Table 3 illustrate the instructional shifts that have occurred because of testing. Two-thirds to three-quarters of all teachers give more emphasis to basic skills instruction, vocabulary lists, word recognition skills, and paper and pencil computation than they would if there were no mandated tests. It should be noted that the pervasiveness of test-directed instructional changes in these high-stakes districts is far greater than those reported by Romberg et al. (1989) where only 30% of teachers reported greater emphasis on basic skills and 25% had increased their emphasis on pencil and paper computation. Our sample differed from the Romberg study in several respects: (1) It focused on districts where testing was known to be high-stakes, while Romberg's nationally representative sample included a variety of high- and low-stakes settings including 13% in which students took no mandated test in mathematics. (2) Our study addressed testing in the elementary grades which we expect to create higher stakes for individual teachers than testing in secondary schools. (3) It is also conceivable that time is a relevant variable, as suggested by the Director of Testing and Research in our largest district who has seen an increasingly frenetic attention to test results despite there having been no formal changes in the district or state testing programs.

The inclusion of the "reading for understanding" item in the skills instruction subscale and the finding that teachers reported increasing this activity in response to testing may seem anomalous if one thinks of this as a high-order, thinking activity. However, we liken this finding to similar results in the Romberg et al. survey where 83% of mathematics teachers said that standardized tests measure problem solving and 30% said they increased their emphasis on problem solving because of mandated testing. Romberg et al. (p. 84) pointed out that teachers appear to consider simple word problems as instances of problem solving rather than adhering to the more ambitious conceptions of problem solving recommended by the NCTM Standards. Although we do not have evidence here of the kinds of activities teachers are using when they emphasize "reading for understanding," it is plausible that they mean activities like those cited by Popham et al. (1985), i.e., finding the main idea and answering questions about passage details in parallel to reading comprehension test questions.

Data in Table 4 show the effects of testing on instructional practices that might have been expected to be slighted if teachers give more emphasis to basic skills. Only the first question, however, clearly followed the predicted pattern. Half of the teachers reported giving less emphasis to "subjects which are not tested" because of standardized testing. For most of the remaining questions the modal response was to say that these various activities were not influenced by standardized testing. Substantial numbers of teachers, however, reported responding to testing by increasing activities such as "kids talking about what's been read" (41%), "extended project work in mathematics" (24%), "reading in books about social studies and science" (42%), "sustained silent reading" (40%) and "work with manipulatives in mathematics" (47%). A clear majority of teachers said that they increased "critical

thinking activities" and "practice in divergent problem solving," 63% and 57% respectively.

When the results in Tables 3 and 4 are taken together it is clear that the great majority of teachers give greater attention to basic skills because of standardized testing. Some teachers follow the predictable pattern of increasing practice on the basics at the expense of more divergent instructional experiences. However, a greater number of teachers appear to increase all activities, especially those pertaining to reading and math instruction. The apparent contradiction, for example, between items 6 and 23, could be explained by saying that teachers give less emphasis to subjects not tested (which includes social studies and science in these districts), but they have students do more reading in books about social studies and science, because this is a reading activity. This interpretation is consistent with the observation of Smith et al. (1990) that with few exceptions, "science at the intermediate grades looks more like reading all the time. Teachers feel they cannot afford to take the time required to set up science activities or do divergent problem-solving. Hence, they spend the time having the pupils read the text and answer the questions at the back of the chapter and take the unit tests (Rottenberg & Smith, 1990, p. 7)." Thus without actual evidence of classroom practices, the claims that critical thinking and divergent problem solving are increased should be interpreted cautiously, especially considering the kinds of instructional practices these same teachers describe in the narrative data in the final section of the paper.

The separate character of the Skills Instruction and Divergent Instruction scales can be seen in their distinct patterns of association with other variables. As shown in Table 13, the Skills scale correlated the most strongly with Test Preparation ($r = .34$) and with Measurement-Driven Instruction ($r = .27$), whereas the correlation of the Divergent scale with these variables was .09 and -.18, respectively. Both Skills and Divergent were correlated with the Pro-tests scale ($r = .20$, and .28). The Divergent Instruction scale correlated negatively ($r = -.40$) with Controversial Testing Practices and negatively with the number of instructional days spent testing ($r = -.27$).

Test preparation

In the preceding section we considered how teachers saw the character of normal instruction changing in response to testing. Test preparation activities also significantly change the instruction children receive in a school year to the extent that test preparation supplants normal instruction. Half of the teachers in these high-stakes districts spend four or more weeks giving students worksheets to review content they expect to be on the test and giving students practice with the kinds of item formats that are on the test. The majority of teachers also reported spending two weeks or more giving students commercially produced test preparation materials, giving practice tests, and instructing students on test-taking strategies. The only test preparation activity that teachers did not report using extensively was giving old standardized tests for practice. The most telling finding, concerning the influence of test preparation on instruction, was that 68% of the teachers reported conducting these test preparation activities "regularly," that is "throughout the school year," rather than limiting them to a few days or weeks before testing.

It should be noted that this was the only question set where there were significant differences between the two districts, with District B reporting more extensive test preparation activities than District A. The two districts did not differ, however, in the pervasive effect of test preparation throughout the school year.

In addition to the previously mentioned correlation between test preparation and instructional emphasis on basic skills, test preparation was also correlated with the Measurement-Driven Instruction scale (r = .26) and with the use of multiple-choice formats in instruction item (r = .26). Test preparation also correlated .42 with increased instructional emphasis on "writing essays" probably because in District B teachers were preparing students to take a mandated writing test.

Controversial testing practices

The section of the teacher questionnaire entitled "Controversial Testing Practices" included a range of test administration practices from those that are clearly cheating to those that might boost scores without being considered unethical. Because of the sensitive nature of these questions teachers were asked to indicate to what extent they were aware of these practices occurring in their schools, rather than being asked to report on their own behavior. Approximately one-quarter of the sample circled the unknown response category.

For most of the controversial testing practices the majority of teachers said that these practices never or rarely happened in their schools. For example, 49% of teachers said that "providing hints on correct answers" rarely or never happened; and "giving students more time than test directions call for" rarely or never happened according to 58% of the teachers. Of course, the negative side to these same data suggest that these kinds of practices do occur to some limited degree. For example, 23% of teachers reported that "providing hints on correct answers" occurred occasionally or frequently. "Rephrasing questions during test administration" was thought to occur occasionally or frequently by 18% of the teachers. Even the two practices which were rejected by the greatest numbers of teachers, "Changing incorrect answers to correct ones on answer documents," and "Encouraging students who would have trouble on the test to be absent," were said to happen occasionally or frequently by 6% and 8% of the teachers respectively.

The most frequently reported controversial testing practices were: "giving practice on highly similar passages," "rephrasing questions during test administration," "providing hints on correct answers," and "giving students more time than test directions call for."

The controversial testing practices scale had some of the strongest correlations with other scales. It correlated .42 with Internal Uses of Tests such as evaluating teachers or evaluating principals, and .24 with External Uses of Tests such as ranking schools in the newspaper. Controversial testing practices also correlated .45 with the Measurement-Driven Instruction scale and .37 with the use of test scores to make student retention and promotion decisions.

Internal and external uses of test data

Responses about the ways that test data are used in the two districts are summarized in Tables 7 and 8. Given that many of the specific questions pertain to district-level practices that should be constant for all teachers within each district, the variation in results must reflect differences in perceptions as much as actual differences in practice. For two of the questions, in fact, the majority of teachers circled the question mark indicating that they did not know whether test scores were used "to allocate or withhold extra district funds" or "to decide how to allocate non-

monetary district resources." Therefore, teachers may not have accurate information about all of the uses of test scores.

Half of the teachers reported that tests are used occasionally or frequently "to compare or evaluate teachers;" 45% said that they are occasionally or frequently used to evaluate principals; and 53% said tests are used occasionally or frequently "to determine awards for school excellence." On the first three items, evaluating teachers, evaluating principals, and allocating district funds there were significant differences between the two districts, with District B reporting these uses more often.

In general external uses of test results were reported to be much more frequent. Tests are used to compare districts, to rank schools in the newspaper, and to compare schools; 76%, 76%, and 71% of teachers said these uses occurred frequently. Teachers reported the use of test scores slightly less frequently for decisions about curriculum and to evaluate innovative programs.

Although constituting distinct factors, the Internal and External Use scales correlated with each other .61. As has already been noted these two scales in turn correlated with controversial testing practices and with the degree of pressure felt to raise test scores.

Positive and negative effects of standardized testing

Questions about the positive and negative effects of standardized testing comprised two scales, Pro Standardized Testing and Measurement-Driven Instruction, which are reported in Tables 9 and 10. Although the items on the Pro Standardized Testing scale are correlated, meaning that respondents who agreed with one item also tended to agree with the other items, there was also a number of respondents "in the middle" so that the majority agreeing or disagreeing switched sides from item to item. Overall teachers rejected more of the pro-testing statements. The items on which a majority either disagreed or strongly disagreed (with corresponding percentages) were "standardized tests help to clarify important learning goals" (65%), "standardized testing is helping schools improve" (64%), "without tests to enforce standards, students would be promoted without prerequisite skills" (71%), "the importance attached to test results gives teachers a sense of common purpose" (72%), "teachers who complain about testing are usually poorer teachers who do not want to be accountable" (77%), and "focusing on tested material first ensures mastery of the basics before going on to other material" (64%).

Only two pro-standardized testing statements received endorsement from a clear majority of teachers: "standardized test results are helpful in identifying student strengths and weaknesses" (72% agreed or strongly agreed) and "low test scores help get additional resources to students with the greatest learning needs" (53% agreed or strongly agreed). Teachers were equally divided on the remaining two statements: "tests give me important feedback about how well I am teaching in each curricular area," and "my school's emphasis on test results shows a real commitment to raising student achievement."

The pattern of results on the Measurement-Driven Instruction scale reported in Table 10 was similar to that for the Pro-Standardized Testing scale. Individual items were strongly intercorrelated meaning that there were many teachers who either consistently endorsed or rejected most of the items; however, there were also a number of teachers in the middle who switched sides thus shifting the majority from item to item.

Four of the Measurement-Driven Instruction items were agreed to by a majority of teachers: "I spend more time teaching reading and math and less time teaching social studies and science because reading and math test scores are so important" (64%), "I use fill-in-the-blank worksheets and matching exercises in my regular instruction so that my students will be comfortable with short answer formats when it comes time to test" (60%), "When I teach reading and math, I emphasize the skills and content I know are on the standardized tests" (69%), and "A lot of the workbook and textbook activities I select for students to do are very similar to the short passages and stand-alone questions that students will encounter on tests" (62%).

Test-driven items with which the majority of teachers disagreed were: "higher-order thinking skills are something I get to only if there is time after covering the basics" (61% disagreed), and "I don't use essay tests during the year because I want my students to have practice with standardized test item formats" (75% disagreed). As noted previously, many teachers in District B reported giving students practice writing essays because they face a mandated writing test.

On the remaining four items, teachers were equally divided as to whether they agreed or disagreed: "My school is more interested in increasing test scores than in improving overall student learning," "Gifted children get to do enrichment activities but at-risk children have to keep drilling on the basics," "There are exciting new curriculum developments that I haven't implemented because they aren't compatible with the tests," and "I tend to drill on basics because that is the only way I can be sure they will get what they really need to know."

Recall that the pro-standardized testing scale was positively correlated with divergent instructional practices. In contrast, the measurement-driven instruction scale was negatively correlated with divergent instruction and positively correlated with pressure, skills instruction, test preparation, controversial testing practices, and both internal and external uses of test results.

Odd items not included in the questionnaire subscales

Three instructional influences items and one test use item did not correlate well with their intended subscales; data for each of these questions are reported separately in Table 11. Writing essays was expected to be a divergent instructional practice while use of multiple choice exercises was expected to go with basic skills instructional emphases. Instead both of these activities were reported to receive greater emphasis as a result of standardized testing, by 71% and 60% of teachers, respectively, and both correlated most highly with the test preparation factor. Giving more emphasis to "timed computational exercises" was reported by 43% of the teachers. This item did not correlated with any scale in the questionnaire, except for a .18 correlation with test preparation.

Forty percent of teachers reported that test scores were occasionally or frequently used to make student promotion or retention decisions. Use of test results in this way positively correlated with test preparation practices ($r = .19$), with controversial testing practices ($r = .37$), with other internal uses of test results ($r = .44$), and with external uses of tests ($r = .26$). Its factor loading was actually higher for the controversial testing practices factor than for the intended test use factor, hence the decision to leave it out of the scales.

Time spent giving standardized tests

Two final questions were asked about the number of hours spent giving standardized tests and the number of days interrupted by testing. Because it had been anticipated that the student testing portion of the larger research project might actually affect teacher questionnaire responses regarding some aspects of testing, the study was designed with teachers selected from two randomly equivalent samples of schools, one sample was involved in testing and the other not. Because there were significant differences between the two samples of teachers on the two questionnaire items pertaining to the amount of time spent testing, data in Table 12 are reported only for the respondents in the nontested schools so that our own testing would not inflate the results. Note that responses for the tested and nontested research groups did not differ significantly on other questionnaire scales.

In District A 58% of the teachers reported spending from 4-8 hours per school year giving standardized tests, thus affecting 3-5 school days; 29% reported spending 9-16 hours testing, with 13% reporting 17 hours or more given to testing. The time given over to testing in District B is clearly greater with 52% saying that 9-16 hours are spent in actual test administrations. Twenty-eight percent of the teachers in District B spend 17 hours or more testing. Not counting test preparation, teachers in District B typically report spending two entire weeks simply giving standardized tests.

Open-ended questions about the positive and negative effects of standardized testing

Teachers were also asked to respond to open-ended questions about the effects of standardized tests. They were asked to give positive examples of how tests helped to improve the quality of education in their classroom or for particular children, and they were asked to give examples of any negative influence of tests on their teaching or student learning. Table 14 provides a lengthy display of teachers' written responses from District B. A comparable analysis was done for District A. There were no apparent differences in types of responses between the two districts, except for the group of respondents in District B who said that the writing test had had a positive effect on their students' writing.

Based on an initial reading of the data, it was decided that the data summary should reflect the overall position of each teacher as well as aggregate the many particular positive and negative examples. Therefore teachers were first sorted into four groups: Nonrespondents (N=47, 17%), Only Positive (N=17, 6%), Both Positive and Negative Answers (N=150, 54%), and Only Negative (N=66, 24%). Then the positive and negative answers were separated and sorted into subcategories. Data within each category were reread and resorted by both authors until the cases were homogeneous within categories and distinguishable between categories. In creating the data table, examples were selected to include both the most typical statements as well as those that were the most different from the category summary. When many teachers gave highly similar answers they were counted but not quoted in the table.

Many teachers gave two or three positive and two or three negative examples. However, within the positive and negative sets only the first example was used to classify a teacher. Therefore the counts for all the subcategories add up to the total for that category. In all but two instances teachers' second and third answers duplicated what had been said by other teachers' first answers; in these two cases teachers said that a negative effect of standardized testing was to "slow down gifted

students." We did not add in the duplicate counts for teachers' second and third answers because it would make it difficult for the reader to make the numbers "add up." However, it should be remembered that many of the high frequency answers both positive and negative would have had even greater endorsements if these counts had been added.

Given all of the national furor about standardized testing, one of the most notable aspects of these data are the many specific positive examples teachers provided of beneficial uses of tests. To be sure there were many negative answers given, and quantitatively they outweighed the positive, but the majority of teachers gave considered, balanced answers. The balance shown argues for the integrity of the data and to a certain extent makes the negative statements all the more compelling because they did not appear to be reflexive, knee-jerk statements against tests.

Six percent of the teachers in District B gave only positive answers and an additional 54% cited positive as well as negative effects. Many of the positive answers are classic "textbook" examples of how test data should be used to diagnose gaps in student knowledge or to evaluate instructional weaknesses. For example, "The areas on which a large % score low (usage, etc.) -- I evaluated and restructured my instruction to improve student understanding and mastery." Other positive answers vindicate reformers intentions when they mandated external tests. For example, "I think the tests encourage teachers to insure learning," and "It helps focus some teachers on basic skills who may otherwise not focus on anything at all."

The decision was made to provide an extensive data table rather than merely a summary of the category labels because this makes more of the data accessible to the reader. A great deal of insightful and meaningful data would be lost if only the summary statements were used. We tried to be faithful to the respondent's view of things in making our classifications. However, if we were to impose our own values on the examples given we might sometimes conclude that positive answers actually reflect negative practices. For example, "Cover more material in less time," "Give me the incentive to drill on a daily basis--for a short, short time--fundamental facts," "Stress on standardized items and repeated repetition helps the slow learners in the group," and so forth. There are also many examples as in the Romberg et al. study where we do not share the same definition of terms as the respondents making it difficult to credit statements such as, "(Standardized testing) enables pupils to do critical thinking." Therefore the detailed information in Table 14 bears careful reading. Depending upon the point of view of the reader, it is possible to see there, even in the positive answers, the negative effects of standardized tests on instruction. Tests set the instructional goals, provide intense motivation, and encourage teachers to address deficiencies skill by skill.

One quarter of all the teachers in District B (24%) gave only negative answers about standardized tests. They either left the question blank which asked for positive examples or said specifically that they "could not think of any." These teachers then have to be combined with the 54% who gave negative examples of test influences along with positive instances. The single largest category of negative response from both groups was the complaint that standardized tests led to "too much teaching to test content and test format." For example, "Critical thinking skills are basically non-existent in our children because of drill and practice for (Test 1), (Test 2), and (Test 3)," "I can't get to science and social studies, like I would like to. Instead, I'm preparing my class for a standardized test," "Too much time is needed to emphasize test content, test taking skills, practice work-sheets," and "We are constantly

reminded to practice, practice, for the test. The fun and excitement has been taken out of teaching." Counting all the responses in this subcategory alone (N = 66 + 28) accounts for 34% of all the teachers who responded from District B, and this does not include teachers who gave answers of this type as their second or third response.

Other small subcategories of negative response add further to the picture of instruction distorted by testing. These include: too much time given to testing and the boring effects of test preparation. These two categories from both negative groups account for an additional 6% of the teacher sample.

The other major categories of negative effects had to do with the stress of testing for both teachers and students and with concerns about the fairness or accuracy of test results. For example, "Too much emphasis and pressure from the principal, district and media. In my opinion, the pressure encourages cheating from a lot of teachers," "Students concentrate on 'what will be included on the test,' students are stressed out and get too nervous during the test." "Our standardized test scores are typed up and passed around to every teacher to compare how they did to others. No mention goes to the teachers with low classes It is much too competitive."

SUMMARY

The purpose of the study was to assess the effects of standardized testing on instruction in two high-stakes school districts. Third, fifth, and sixth grade teachers in approximately 100 schools were surveyed with a questionnaire. The study was intended specifically to examine testing and instruction practices in settings where testing is high stakes. Findings clearly cannot be generalized to other contexts.

A limitation of the study was the 42% response rate. The 360 teachers who responded are not necessarily representative of all teachers even in the high-stakes districts. For respondents, however, there were a number of indicators that teachers answered carefully and took the task of filling out the questionnaire seriously. Thus we judged the quality or integrity of the data to be good. Only 20% did not complete the open-ended questions. On several scales the majority opinion shifted sides often suggesting that respondents did not simply go down the page answering strongly agree or strongly disagree to all items. There were few questions left blank except for those questions where we had anticipated that teachers might not know the answer and had provided a question mark category. The factor analysis showed highly internally consistent scales and nuanced distinctions between scales, which would not have been possible with careless responses. Answers to factual questions such as the amount of time spent testing were consistent with what we were told by the Directors of Research and Testing in each district.

The major findings of the study were as follows:

1. Teachers reported that they feel pressured to improve test scores by the district administration and by the media. 79% of teachers said that they feel "substantial" or "great" pressure from the district administration to raise test scores; 66% said they felt such pressure from the newspapers and media. Within these high-stakes districts, perceived pressure to raise test scores was correlated with measurement-driven instructional practices, the use of tests to evaluate teachers and principals, and the extent of test preparation practices.

2. Because of the importance of standardized tests, teachers give greater emphasis to basic skills instruction. Two-thirds to three-quarters of teachers reported giving more emphasis to basic skills instruction, vocabulary lists, word recognition skills, and paper and pencil computation than they would if there were no mandated tests. Teachers also acknowledged the influence of the tests on how they taught the basics. For example, 69% agreed with the statement, "When I teach reading and math, I emphasize the skills and content I know are on the standardized tests." In the narrative data from the open-ended questions, the single largest category of negative response was the complaint that standardized tests lead to "too much teaching to test content and test format."

3. Non-tested content clearly suffers because of the focus on standardized tests. Half of the teachers said that they give less emphasis to "subjects which are not tested." In the open-ended questions, science and social studies were explicitly identified as the subjects that are slighted. Many teachers also said that they give less attention to higher-order thinking and extended projects because of testing. However, a different group of teachers said that they increased these activities (especially in reading and math) in response to testing.

4. In addition to the effect of testing on the character of normal instruction, testing also distorts instruction because of the extensive time given to test preparation. Half of the teachers in these high-stakes districts spend four or more weeks per year giving students worksheets to review content they expect to be on the test and giving students practice with the kinds of item formats that are on the test. A majority of teachers also reported spending two weeks or more giving students commercially produced test preparation materials, giving practice tests, and instructing students on test-taking strategies. The most telling finding, concerning the influence of test preparation on instruction, was that 68% of the teachers reported conducting these test preparation activities "regularly," that is "throughout the school year," rather than limiting them to a few days or weeks before testing.

5. Four weeks of test preparation does not count the one or two full weeks of school spent giving tests. In District A the modal response for number of school days interrupted by standardized testing was 3-5 days. In District B it was 6-10 days, with 29% of District B's teachers saying that they spent 11 or more days giving tests.

6. Teachers reported that flagrant instances of cheating happen very rarely in their schools but other controversial testing practices that would clearly boost scores happen more frequently. For example, the two most extreme practices, changing answer documents or encouraging low scorers to be absent, were reported by only 6% and 8% of teachers respectively. However, other practices such as "rephrasing questions during test administration," "providing hints on correct answers," and "giving students more time than test directions call for," were thought to happen occasionally or frequently by 18%, 23%, and 20% of the teachers respectively.

The controversial testing practices scale correlated .42 with internal uses of tests such as evaluating teachers or principals and .45 with the measurement-driven instruction scale.

7. Teachers reported extensive use of test results for external purposes such as comparing districts, ranking schools in the newspaper, and comparing schools. They reported only slightly less use of tests for internal purposes such as evaluating principals and evaluating teachers. The more tests were used for these purposes,

the higher was the report of controversial testing practices and measurement-driven instruction.

8. In open-ended questions, the majority of teachers could identify numerous beneficial uses of standardized tests, such as clearly setting instructional goals, providing feedback about student strengths and weaknesses, and identifying gaps in instruction. For all but a small fraction of teachers, however, these benefits from standardized testing were offset or greatly outweighed by negative effects, such as the amount of instructional time given to test preparation, the amount of stress experienced, unfair or invalid comparisons, and the demoralizing effects on teachers and students. One-quarter of the teacher sample reported only negative influences of standardized tests on teaching and learning.

REFERENCES

Bracey, G.W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delta Kappan, 68, 683-686.

Cannell, J.J. (1987). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. Daniels, WV: Friends for Education.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. The Elementary School Journal, 85, 315-336.

Goslin, D.A. (1967). Teachers and testing. New York: Russell Sage Foundation

Koretz, D.M., Linn, R.L., Dunbar, S.B., & Shepard, L.A. (1991, April). The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests. Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

National Council of Teachers of Mathematics. (1989). **Curriculum and evaluation standards for school mathematics.** Reston, VA: Author.

Popham, W.J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68, 679-682.

Popham, W.J., Cruse, K.L., Rankin, S.C., Sandifer, R.D., & Willaims, P.L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66, 628-634.

Romberg, T.A., Zarinnia, E.A., & Williams, S.R. (1989, March). The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions. Madison: National Center for Research in Mathematical Science Education, University of Wisconsin-Madison.

Rottenberg, C., & Smith, M.L. (1990, April). Unintended effects of external testing in elementary schools. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Shepard, L.A. (1990a). Inflated test score gains: Is the problem old norms or teaching the test? Educational Measurement: Issues and Practice, 9, 15-22.

Shepard, L.A. (1990b, April). Psychometricians' beliefs about learning. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Smith, M.L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1990). The role of testing in elementary schools. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

Table 1
Factor Analysis of the Teacher Questionnaire:
Oblique Rotation Pattern Matrix with Nine Factors Extracted (N=120-350)

	FACTOR 1 (controv)	FACTOR 2 (skillins)	FACTOR 3 (testprep)	FACTOR 4 (extuse)	FACTOR 5 (pro-tests)	FACTOR 6 (measdriv)	FACTOR 7 (divergen)	FACTOR 8 (pressure)	FACTOR 9 (intuse)
CONTR38	.87	.02	-.04	.01	-.05	.04	.09	-.03	.01
CONTR45	.84	.04	.06	-.09	.05	.05	-.10	.05	.02
CONTR36	.81	.03	.02	.03	.11	-.01	-.07	-.03	-.04
CONTR44	.80	.03	.01	-.01	.07	-.03	-.08	.00	.03
CONTR37	.80	.07	-.04	.01	.00	.04	-.04	-.02	-.05
CONTR39	.77	.03	.10	.06	-.10	-.30	-.20	-.15	.05
CONTR40	.72	.10	-.09	-.09	.00	.03	-.07	.05	.08
CONTR41	.70	.08	.04	.13	-.06	-.07	-.03	-.02	.04
CONTR42	.65	-.06	-.02	-.16	-.02	-.04	-.08	.08	-.04
CONTR46	.58	-.04	.03	.09	.09	-.14	-.03	.07	-.09
CONTR43	.57	.01	.00	-.02	.06	-.01	-.03	.09	-.09
USE47	.26	-.02	-.12	.10	-.08	.03	.06	-.04	-.22
INST9	.10	.85	.09	.06	.04	-.01	-.03	.00	.10
INST10	.06	.83	.04	.05	.00	.00	-.04	.04	.09
INST11	-.01	.73	.00	.02	.02	-.05	.08	-.02	.01
INST13	.11	.67	-.04	.01	.01	.08	.10	.05	.02
INST19	-.02	.46	-.02	.05	-.05	-.25	.05	.03	-.03
INST12	.07	.44	-.13	-.03	.10	.03	.07	.00	-.06
INST20	-.06	.30	-.19	-.19	.04	-.20	-.01	.01	-.08
PREP30	.00	-.05	-.87	.13	-.06	-.05	.07	.05	.10
PREP31	-.02	-.06	-.81	-.01	-.02	-.08	-.04	.09	.10
PREP34	-.04	.03	-.77	.00	-.07	.10	.14	.01	.00
PREP29	.00	-.10	-.75	.08	-.02	-.17	-.01	.11	.07
PREP32	.00	-.02	-.74	-.01	.04	-.08	-.04	.09	-.05
PREP33	.11	.01	-.60	-.19	.04	.00	.00	.00	-.11
PREP35	-.01	.11	-.41	.07	.06	.10	.00	-.10	-.09
INST7	-.03	-.10	-.35	-.08	.07	.07	.03	.05	-.19
INST8	.02	.15	-.20	-.02	-.04	-.08	-.02	.00	-.04
USE51	.04	-.04	.06	.78	-.07	-.13	.00	-.02	-.07
USE50	.07	-.04	.09	.76	-.09	-.19	.05	.02	-.12
USE54	-.06	.10	.09	.69	-.07	-.09	-.04	.09	-.05
USE52	.00	.07	-.10	.64	.09	.13	-.06	.09	-.07
USE53	-.03	.01	-.21	.56	.10	.09	.00	-.06	-.10
EFFECT62	-.03	.02	-.02	-.03	.65	.05	.09	-.01	-.03
EFFECT68	.06	-.08	.00	-.01	.64	.03	.04	.00	-.03
EFFECT61	-.03	.05	.06	-.07	.63	-.13	.00	-.03	-.02
EFFECT71	.06	.06	.04	.07	.62	-.01	-.03	-.01	.01
EFFECT75	.06	.01	.00	-.03	.58	-.20	.02	-.05	-.08
EFFECT72	-.17	.06	-.16	.06	.54	.13	-.03	-.01	.07
EFFECT67	.02	.03	.01	.07	.50	-.05	.06	.03	.03
EFFECT74	.10	-.11	.00	-.07	.44	-.07	.03	-.09	-.15
EFFECT58	-.06	.00	.06	.03	.43	.01	-.03	.13	.12
EFFECT59	.00	.10	-.05	-.10	.30	.10	.05	-.08	.05
EFFECT76	.12	-.03	.00	.00	-.03	-.60	.01	.10	.00
EFFECT70	.02	.05	-.02	.03	-.02	-.56	-.10	-.12	-.06
EFFECT77	-.04	-.04	.02	.03	.19	-.53	.08	.13	.04
EFFECT65	.06	.06	.08	.01	-.02	-.52	-.17	-.02	.06
EFFECT63	.25	-.07	.00	.00	-.23	-.46	.01	.15	-.03
EFFECT66	.14	.07	-.18	.11	.07	-.46	-.01	.09	.04
EFFECT64	-.04	.04	-.12	-.04	.00	-.46	.00	-.13	-.03
EFFECT60	.10	.15	-.04	-.01	.06	-.45	-.01	.14	.02
EFFECT69	.07	.04	.12	.10	.02	-.45	-.12	.08	-.05
EFFECT73	.00	-.02	-.21	-.04	.19	-.41	.04	-.02	-.06
INST26	-.07	-.05	.00	.01	.03	-.08	.65	-.10	-.03
INST28	-.01	.21	-.08	-.09	.02	.08	.58	-.07	.06
INST15	-.07	-.10	-.03	-.09	.03	.08	.58	.00	.02
INST21	-.15	.16	.00	-.16	-.04	.04	.56	.12	-.07
INST23	-.03	.06	-.08	.00	.00	-.02	.53	-.12	-.03
INST27	.02	.28	-.09	-.06	-.02	.06	.51	-.06	.06
INST25	-.20	.21	-.06	-.04	.08	.03	.50	.06	-.10
INST22	-.14	.21	-.02	-.08	.12	.08	.47	.02	-.13
INST18	.05	-.11	.05	.14	.11	.09	.46	-.02	.13
INST24	-.05	.18	.07	.02	.07	-.05	.41	-.02	-.09
INST17	.05	-.17	.09	.01	.04	.05	.40	-.06	.08
INST14	-.19	.32	.07	-.04	-.03	.00	.40	.08	-.09
INST6	-.06	-.19	.10	.01	.09	.16	.30	-.04	.01
INST16	.03	.03	-.13	.08	-.02	-.10	.19	.03	.08
PRESS3	.02	-.05	-.14	.10	-.09	-.03	-.06	.61	.03
PRESS5	-.02	.04	-.14	.15	-.03	.14	-.13	.60	-.04
PRESS2	-.02	.09	-.08	-.13	.07	-.10	-.06	.59	-.07
PRESS1	.16	.03	-.05	-.08	-.07	-.22	.02	.52	-.02
PRESS4	.00	-.01	.06	.02	.02	.08	.05	.47	.00
USE55	.00	-.06	-.02	.16	.01	-.03	-.05	.05	-.84
USE56	.00	-.02	-.02	.25	.02	.04	-.10	-.01	-.81
USE48	.25	-.08	.00	.14	-.05	-.14	.00	.20	-.42
USE57	.07	-.05	.15	.25	-.03	.05	.05	.16	-.39
USE49	.27	-.09	-.07	.24	-.11	-.11	-.02	.13	-.33

Table 2

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: PRESSURE

PRESSURE FOR IMPROVED TEST SCORES. To what extent do you feel pressure from the following groups to improve your students' standardized test scores?

No Pressure	Slight Pressure	Moderate Pressure	Substantial Pressure	Great Pressure	*Place a check in the appropriate column to show how much pressure to improve test scores you personally feel from each of the groups.*
					1. My principal
					2. Other teachers
					3. District administration or board of education
					4. Parents
					5. Newspaper/media

	Frequencies in %						Means		
#	no press (0)	slight press (1)	mod. press (2)	subst. press (3)	great press (4)	blank	District A	District B	p<.001
1	5.5	11.1	25.2	26.9	29.4	1.9	3.50	3.69	
2	26.3	24.7	20.8	19.7	4.7	3.9	1.93	2.67	
3	2.2	5.5	10.8	26.3	52.6	2.5	4.10	4.29	
4	37.1	20.5	21.9	11.6	5.5	3.3	2.04	2.32	
5	9.1	7.5	15.5	19.1	46.8	1.9	3.48	4.01	

	Combined	District A	District B
Subscale Total Score:	16.42 (sd=4.22)	15.04(4.02)	16.95(4.12)
Average Item Score:	3.3	3.0	3.4
N=	350	80	262

Table 3

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: SKILLS INSTRUCTION (SKILLINS)

TESTING AND INSTRUCTION. Items 6-28 describe a variety of instructional activities. Circle the appropriate response to show whether the use of standardized tests in your school causes you to place MORE EMPHASIS or LESS EMPHASIS on the activity than if there were no mandated standardized tests, or whether your emphasis on the activity is NOT INFLUENCED by standardized testing.

9. Basic skills in mathematics ____ More Emphasis ____ Not Influenced ____ Less Emphasis
10. Basic skills in reading ____ More Emphasis ____ Not Influenced ____ Less Emphasis
11. Reading for understanding ____ More Emphasis ____ Not Influenced ____ Less Emphasis
12. Vocabulary lists ____ More Emphasis ____ Not Influenced ____ Less Emphasis
13. Word recognition skills ____ More Emphasis ____ Not Influenced ____ Less Emphasis
19. Paper and pencil computation ____ More Emphasis ____ Not Influenced ____ Less Emphasis
20. Direct instruction to the whole class ____ More Emphasis ____ Not Influenced ____ Less Emphasis

	Frequencies in %				Means		
#	more emphasis (3)	not influenced (2)	less emphasis (1)	blank	District A	District B	p<.001
9	74.0	24.9	0.0	1.1	2.75	2.75	
10	75.3	23.5	0.6	0.6	2.70	2.77	
11	72.9	25.8	0.8	0.6	2.64	2.75	
12	66.2	31.3	2.2	0.3	2.46	2.69	*
13	69.0	27.7	2.8	0.6	2.58	2.69	
19	65.7	29.9	2.2	2.2	2.54	2.68	
20	53.5	42.1	3.0	1.4	2.39	2.55	

	Combined	District A	District B
Subscale Total Score:	18.69 (sd=2.46)	18.05(2.57)	18.93(2.37)
Average Item Score:	2.7	2.6	2.7
N=	354	77	269

Table 4

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: DIVERGENT INSTRUCTIONAL PRACTICES (DIVERGEN)

6. Subjects which are not tested ____ More Emphasis ____ Not Influenced ____ Less Emphasis
14. Kids talking about what's been read ____ More Emphasis ____ Not Influenced ____ Less Emphasis
15. Extended project work in mathematics ____ More Emphasis ____ Not Influenced ____ Less Emphasis
17. Activities involving calculators ____ More Emphasis ____ Not Influenced ____ Less Emphasis
18. Activities involving computers ____ More Emphasis ____ Not Influenced ____ Less Emphasis
21. Small group instruction ____ More Emphasis ____ Not Influenced ____ Less Emphasis
22. Cooperative learning ____ More Emphasis ____ Not Influenced ____ Less Emphasis
23. Reading in books about social studies and science ____ More Emphasis ____ Not Influenced ____ Less Emphasis
24. Sustained silent reading ____ More Emphasis ____ Not Influenced ____ Less Emphasis
25. Work with manipulatives in mathematics ____ More Emphasis ____ Not Influenced ____ Less Emphasis
26. Library projects and report writing ____ More Emphasis ____ Not Influenced ____ Less Emphasis
27. Critical thinking activities ____ More Emphasis ____ Not Influenced ____ Less Emphasis
28. Practice in divergent problem solving ____ More Emphasis ____ Not Influenced ____ Less Emphasis

	Frequencies in %				Means		
#	more emphasis (3)	not influenced (2)	less emphasis (1)	blank	District A	District B	p<.001
6	10.2	38.2	49.6	1.9	1.76	1.55	
14	41.0	47.1	11.4	0.6	2.29	2.30	
15	24.1	49.6	24.9	1.4	2.10	1.96	
17	5.5	59.0	33.2	2.2	1.79	1.69	
18	13.9	59.3	24.7	2.2	2.06	1.84	
21	34.9	51.0	13.3	0.8	2.23	2.22	
22	39.3	47.6	12.2	0.8	2.24	2.28	
23	42.4	41.8	14.7	1.1	2.28	2.28	
24	39.9	46.3	13.0	0.8	2.20	2.29	
25	46.5	40.2	11.9	1.4	2.39	2.34	
26	27.1	46.8	24.9	1.1	2.15	1.99	
27	62.9	26.0	10.0	1.1	2.49	2.55	
28	57.3	30.5	11.1	1.1	2.46	2.47	

	Combined	District A	District B
Subscale Total Score:	28.01 (sd=5.18)	28.51(4.15)	27.78(5.45)
Average Item Score:	2.2	2.2	2.1
N=	344	78	258

Table 5

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: TEST PREPARATION (TESTPREP)

TEST PREPARATION ACTIVITIES. For items 29-35, try to recall how much time over this entire school year you spent in your classroom on the following test preparation activities. Then mark one box for each item.

29. Giving students worksheets that review the content you expect to be on the test
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
30. Giving students practice with the kinds of item formats that are on the test
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
31. Giving students commercially produced test preparation materials
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
32. Giving students practice tests developed by school, district or state staff
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
33. Giving students old standardized test forms for practice
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
34. Instructing students on test-taking strategies
    [ ] No time [ ] A day or less [ ] 2-5 days [ ] 2-3 weeks [ ] 4 or more weeks
35. When did most of the test preparation activities you conducted take place?
    [ ] A few days before the testing [ ] Regularly throughout the school year
    [ ] A few weeks before the testing [ ] Not applicable, we didn't do any test preparation.

	Frequencies in %						Means		
#	no time (0)	a day or less (1)	2-5 days (2)	2-3 weeks (3)	4+ weeks (4)	blank	District A	District B	p<.001
29	6.9	3.6	15.2	21.9	51.5	0.8	3.29	4.31	*
30	2.5	5.8	15.8	24.4	50.7	0.8	3.49	4.35	*
31	12.7	6.6	18.8	25.5	35.5	0.8	2.70	3.92	*
32	7.5	9.4	21.9	21.3	38.5	1.4	2.74	4.04	*
33	44.6	5.0	17.5	13.9	18.6	0.6	1.56	2.85	*
34	0.8	13.0	22.4	19.1	43.8	0.8	3.27	4.11	*

	days before (1)	weeks before (2)	regularly (3)	not applic.	blank			
35	5.8	23.3	68.1	0.3	2.5		2.53	2.68

	Combined	District A	District B
Subscale Total Score:	24.69 (sd=6.32)	16.25(5.86)	26.37(5.40)
Average Item Score (items 29-34):	3.6	2.7	4.4
N=	350	78	265

Table 6

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: CONTROVERSIAL TESTING PRACTICES (CONTROV)

CONTROVERSIAL TESTING PRACTICES. Items 36-46 list testing practices some teachers use to improve students' test scores. To what extent do you believe they are practiced by teachers in your school? Use this 5-point scale to tell us the extent to which each of the practices listed in items 36 to 46 occurs in your school; simply circle the code that matches the frequency of each practice.

N = NEVER happens in my school
R = RARELY happens in my school
O = OCCASIONALLY happens in my school
F = FREQUENTLY happens in my school
? = NO IDEA how frequently the practice occurs

36. Providing hints on correct answers.....N.....R.....O.....F.....?
37. Giving students more time than test directions call for.....N.....R.....O.....F.....?
38. Reading questions that students are supposed to read themselves.....N.....R.....O.....F.....?
39. Answering questions during testing time about test content.....N.....R.....O.....F.....?
40. Changing incorrect answers to correct ones on answer documents.....N.....R.....O.....F.....?
41. Rephrasing questions during test administration.....N.....R.....O.....F.....?
42. Not administering the test to students who would have trouble on the test.....N.....R.....O.....F.....?
43. Encouraging students who would have trouble on the test to be absent.....N.....R.....O.....F.....?
44. Practicing on items from the test itself.....N.....R.....O.....F.....?
45. Giving students answers to test questions.....N.....R.....O.....F.....?
46. Giving practice on highly similar passages.....N.....R.....O.....F.....?

	Frequencies in %					Means		
#	N (1)	R (2)	O (3)	F (4)	? [a]	District A	District B	p<.001
36	28.5	20.8	16.9	5.8	28.0	1.91	2.02	
37	38.0	19.7	15.2	4.4	22.7	1.71	1.85	
38	38.8	22.2	11.9	2.2	24.9	1.77	1.68	
39	43.2	20.5	8.9	2.8	24.7	1.70	1.56	
40	58.4	7.8	5.5	0.6	27.7	1.16	1.31	
41	36.3	20.8	16.1	1.9	24.9	1.77	1.79	
42	50.7	15.8	7.5	5.8	20.2	1.29	1.68	
43	60.1	10.8	5.5	1.9	21.6	1.10	1.42	
44	54.6	12.5	8.0	3.3	21.6	1.46	1.50	
45	56.8	11.6	6.4	1.9	23.3	1.33	1.41	
46	24.9	15.8	20.5	19.7	19.1	2.14	2.50	

Subscale Total Score:	Combined	District A	District B
	17.17 (sd=6.94)	16.25(5.86)	17.57(7.26)
Average Item Score:	1.6	1.5	1.6
N=	207		

[a]Note: The question mark frequencies include the small percentage of teachers who left the question blank as well as those those who circled the ? to indicate that they could not answer. Only response categories 1-4 were used in the calculation of means, hence the dramatic reduction in sample size compared to other tables.

Table 7

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: INTERNAL USES OF TESTS (INTUSE)

USES OF TEST DATA. Items 47-57 list some uses of standardized test scores. Use the same 5-point scale to indicate how frequently standardized test scores are used in your district for the following purposes.

48. To compare or evaluate teachers.....N.....R.....O.....F.....?
49. To compare or evaluate principals.....N.....R.....O.....F.....?
55. To allocate or withhold extra district funds .....N.....R.....O.....F.....?
56. To decide how to allocate non-monetary district resources.....N.....R.....O.....F.....?
57. To determine awards for school excellence .....N.....R.....O.....F.....?

	Frequencies in %					Means		
#	N (1)	R (2)	O (3)	F (4)	? [a]	District A	District B	p<.001
48	14.1	14.4	22.4	28.0	21.1	2.37	2.93	*
49	12.5	10.2	16.1	29.1	32.1	2.26	3.07	*
55	10.8	8.6	10.0	21.1	49.6	2.27	2.98	*
56	8.3	7.2	11.6	20.2	52.7	2.49	3.04	
57	8.3	4.4	15.8	37.4	34.1	3.39	3.21	

	Combined	District A	District B
Subscale Total Score:	14.15 (sd=4.55)	12.28(4.24)	14.89(4.36)
Average Item Score:	2.8	2.5	3.0
N=	143	29	110

[a]Note: The question mark frequencies include the small percentage of teachers who left the question blank as well as those who circled the ? to indicate that they could not answer. Only response categories 1-4 were used in the calculation of means, hence the dramatic reduction in sample size compared to other tables.

Table 8

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: EXTERNAL USES OF TESTS (EXTUSE)

USES OF TEST DATA. Items 47-57 list some uses of standardized test scores. Use the same 5-point scale to indicate how frequently standardized test scores are used in your district for the following purposes.

50. To compare district schools against one another ..... N ..... R ..... O ..... F ..... ?
51. To compare district performance against other districts ..... N ..... R ..... O ..... F ..... ?
52. To make decisions about what curriculum to emphasize ..... N ..... R ..... O ..... F ..... ?
53. To decide on continuation of innovative programs ..... N ..... R ..... O ..... F ..... ?
54. To rank schools in the newspaper ..... N ..... R ..... O ..... F ..... ?

	Frequencies in %					Means		
#	N (1)	R (2)	O (3)	F (4)	? [a]	District A	District B	p<.001
50	3.0	3.9	13.6	71.2	8.3	3.55	3.70	
51	1.9	3.0	12.7	75.9	6.4	3.61	3.77	
52	1.4	3.3	16.6	63.2	15.5	3.53	3.71	
53	3.6	7.5	19.9	42.9	26.0	3.29	3.40	
54	2.8	2.5	10.0	75.9	8.9	3.61	3.78	

	Combined	District A	District B
Subscale Total Score:	18.15 (sd=2.81)	17.73(2.76)	18.27(2.80)
Average Item Score:	3.6	3.5	3.7
N=	247	44	197

[a]Note: The question mark frequencies include the small percentage of teachers who left the question blank as well as those who circled the ? to indicate that they could not answer. Only response categories 1-4 were used in the calculation of means, hence the dramatic reduction in sample size compared to other tables.

Table 9

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: PRO STANDARDIZED TESTING (PRO-TESTS)

THE EFFECTS OF STANDARDIZED TESTING. Today in the United States there is intense debate about whether the standardized testing imposed on schools for accountability purposes is good or bad. A number of potentially positive and negative effects of standardized tests are listed in items 58-77.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	No Opinion	*Place a check in the appropriate column to show how much you personally agree with each of these statements.*
					58. Standardized test results are helpful in identifying student strengths and weaknesses.
					59. Low test scores help get additional resources to students with the greatest learning needs.
					61. Standardized tests help to clarify which learning goals are the most important.
					62. Standardized testing is helping schools improve.
					67. Without tests to enforce standards, students would move to the next grade without prerequisite skills.
					68. Tests give me important feedback about how well I am teaching in each curricular area.
					71. The importance attached to test results gives teachers a sense of common purpose.
					72. My school's emphasis on test results shows a real commitment to raising student achievement.
					74. Teachers who complain about testing are usually poorer teachers who do not want to be accountable.
					75. Focusing on tested material first ensures mastery of the basics before going on to other material.

	Frequencies in %					Means		
#	strongly agree (4)	moderat. agree (3)	moderat. disagree (2)	strongly disagree (1)	no opinion [a]	District A	District B	p<.001
58	21.3	60.4	10.5	5.5	2.2	2.93	3.01	
59	12.7	39.9	17.5	21.9	8.0	2.60	2.44	
61	7.5	19.7	29.4	35.2	8.3	1.96	2.00	
62	4.4	25.2	26.0	38.0	6.4	1.89	1.98	
67	4.7	18.8	24.1	47.1	5.3	1.69	1.83	
68	9.1	39.1	23.0	24.7	4.2	2.31	2.35	
71	4.2	16.9	22.7	49.3	6.9	1.66	1.76	
72	11.9	31.3	22.7	24.4	9.7	2.24	2.37	
74	3.9	10.0	20.5	56.8	8.9	1.25	1.67	
75	6.1	24.4	24.7	39.3	5.5	1.85	2.00	

Subscale Total Score:	Combined	District A	District B
	20.75 (sd=5.54)	19.84(4.78)	21.01(5.73)
Average Item Score:	2.1	2.0	2.1
N=	251	51	193

[a]Note: The question mark frequencies include the small percentage of teachers who left the question blank as well as those who circled the ? to indicate that they could not answer. Only response categories 1-4 were used in the calculation of means, hence the dramatic reduction in sample size compared to other tables.

Table 10

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON THE SUBSCALE: MEASUREMENT-DRIVEN INSTRUCTION (MEASDRIV)

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	No Opinion	*Place a check in the appropriate column to show how much you personally agree with each of these statements.*
					60. I spend more time teaching reading and math and less time teaching social studies and science because reading and math test scores are so important.
					63. My school is more interested in increasing test scores than in improving overall student learning.
					64. I use fill-in-the-blank worksheets and matching exercises in my regular instruction so that my students will be comfortable with short answer formats when it comes time to test.
					65. Higher-order thinking skills are something I get to only if there is time after covering the basics.
					66. When I teach reading and math, I emphasize the skills and content I know are on the standardized tests.
					69. Gifted children get to do enrichment activities but at-risk children have to keep drilling on the basics.
					70. I don't use essay tests during the your because I want my students to have practice with standardized test item formats.
					73. A lot of the workbook and textbook activities I select for students to do are very similar to the short passages and stand-alone questions that students will encounter on tests.
					76. There are exciting new curriculum developments (e.g., whole language, minds-on-science) that I haven't implemented because they aren't compatible with the tests we're measured by.
					77. I tend to drill students on basic skills because that is the only way I can be sure they will get what they really need to know.

	Frequencies in %					Means		
#	strongly agree (4)	moderat. agree (3)	moderat. disagree (2)	strongly disagree (1)	no opinion [a]	District A	District B	p<.001
60	34.1	29.9	14.4	17.5	4.2	2.68	2.88	
63	22.4	23.5	24.4	24.7	5.0	2.24	2.52	
64	19.9	39.9	19.4	15.0	5.8	2.81	2.65	
65	12.7	24.4	20.5	40.2	2.2	2.30	2.04	
66	31.9	36.8	14.4	14.1	2.8	2.51	2.99	
69	19.4	28.0	18.3	28.0	6.4	2.30	2.45	
70	5.5	9.7	25.8	49.0	10.0	1.66	1.69	
73	18.6	43.5	18.6	14.4	5.0	2.58	2.73	
76	15.5	26.6	16.9	28.0	13.0	2.11	2.40	
77	10.5	36.3	22.4	24.4	6.4	2.31	2.36	

	Combined	District A	District B
Subscale Total Score:	24.40 (sd=6.46)	23.24(6.38)	24.93(6.40)
Average Item Score:	2.4	2.3	2.5
N=	244	51	187

[a]Note: 'No opinion' responses are excluded from the calculation of means.

Table 11

ITEM MEANS AND FREQUENCIES AND SUBTEST STATISTICS FOR TWO HIGH-STAKES DISTRICTS ON ODD ITEMS: WRITING ESSAYS, MULTIPLE CHOICE EXERCISES, COMPUTATION EXERCISES, RETENTION DECISIONS

7. Writing essays _____ More Emphasis _____ Not Influenced _____ Less Emphasis
8. Use of multiple choice, fill-in and matching exercises _____ More Emphasis _____ Not Influenced _____ Less Emphasis
16. Timed computational exercises _____ More Emphasis _____ Not Influenced _____ Less Emphasis

47. To promote or retain students.....N.....R.....O.....F.....?

Frequencies in %

#	more emphasis (3)	not influenced (2)	less emphasis (1)	blank
7	71.2	23.0	5.5	0.3
8	60.4	33.8	3.9	1.9
16	43.2	47.1	8.6	1.1

	Combined	District A	District B	
Average Item Score:# 7	2.65 (sd=0.58)	2.04	2.84	p<.001
N=	369	80	280	
	Combined	District A	District B	
Average Item Score:# 8	2.58 (sd=0.57)	2.54(0.59)	2.59(0.56)	
N=	354	80	274	
	Combined	District A	District B	
Average Item Score:# 16	2.35 (sd=0.63)	2.29	2.37	
N=	366			

Frequencies in %

#	N (1)	R (2)	O (3)	F (4)	?[a]
47	25.5	22.7	26.6	13.9	11.4

	Combined	District A	District B	
Average Item Score:	2.33 (sd=1.06)	2.10(0.94)	2.39(1.08)	p<.001
N=	328	71	249	

[a]Note: The question mark frequencies include the small percentage of teachers who left the question blank as well as those who circled the ? to indicate that they could not answer. Only response categories 1-4 were used in the calculation of means, hence the dramatic reduction in sample size compared to other tables.

Table 12
FREQUENCIES FOR HOURS AND DAYS SPENT GIVING STANDARDIZED TESTS

80. Estimate how much time you have spent in your classroom this school year giving standardized tests. (Administering the reading and math subtests of a typical standardized test takes about 4-5 hours.)
[ ] Less than 4 hours [ ] 4-8 hours [ ] 9-16 hours [ ] 17 or more hours

81. How many days have been interrupted this school year as a result of giving standardized tests?
[ ] 2 or fewer [ ] 3-5 [ ] 6-10 [ ] 11 or more

#		<4 hours	4-8 hours	9-16 hours	>17 hours	blank
80	District A N=24	0.0	58.3	29.2	12.5	0.0
	District B N=99	3.0	16.2	51.5	28.3	1.0
	Combined N=123	2.4	24.4	47.2	25.2	.8

#		2/fewer days	3-5 days	6-10 days	11+ days	blank
81	District A N=24	0.0	79.2	16.7	4.2	0.0
	District B N=100	3.0	12.0	56.0	29.0	0.0
	Combined N=124	2.4	25.0	48.4	24.2	0.0

Table 13
CORRELATIONS AMONG SUBSCALES AND ODD ITEMS

	Pressure	Skillins	Divergen	Testprep	Controv	Intuse	Extuse	Protests	Measdriv	Essays	Multip	Timecom	Promretn	Testhrs	Intdays
Pressure	.73														
Skillins	**.13**	.83													
Divergen	**-.18**	**.26**	.84												
Testprep	**.24**	**.34**	.09	.87											
Controv	**.23**	.10	**-.40**	.06	.94										
Intuse	**.29**	-.10	**-.24**	.03	**.42**	.84									
Exuse	**.22**	-.07	-.12	.00	**.24**	**.61**	.84								
Protest	**-.16**	**.20**	**.28**	.12	-.16	-.08	-.09	.81							
Measdriv	**.30**	**.27**	**-.18**	**.26**	**.45**	.21	**.21**	.10	.82						
Essays	.07	**.23**	**.14**	**.42**	-.08	.13	-.02	.14	.04	.87					
Multip	.09	**.23**	.00	**.26**	.09	.02	-.03	.01	.10	.06	1.00				
Timecomp	.06	.16	.14	**.18**	.05	.03	.06	.02	.14	.00	.09	1.00			
Promretn	.10	.05	-.10	**.19**	**.37**	**.44**	**.26**	-.11	.14	.07	.08	.05	1.00		
Testhrs	**.14**	.03	-.12	**.22**	**.19**	.00	-.04	-.13	.15	.04	.01	-.05	.12	1.00	
Intdays	**.20**	.00	**-.27**	**.21**	**.20**	.16	.08	**-.18**	**.20**	**.14**	-.02	-.08	**.14**	**.47**	1.00

NOTE: Internal consistency coefficients are reported on the diagonal. Ns for individual pairwise comparisons ranged from 102 (for the MEASDRIV by INTUSE correlation) to 346.
Correlations significant at $p < .01$ are in bold face type.

Table 14
Summary of Teachers' Answers to Open-ended Questions on the Positive and Negative Effects of Standardized Testing
(District B, N = 280)

Personal Experience with Standardized Tests. The debate about testing--pro and con--is often vague. Do you have specific examples of how standardized tests have had an impact on you or your students?

78. Give one or two examples of how standardized tests helped you to improve the quality of education in your classroom or for particular children.

79. Give one or two examples of how standardized tests had a negative influence on your teaching or student learning.

Non Response: 17% (N=47) did not respond to the open-ended questions.

Only Positive: 6% (N=17) gave only positive answers, some said specifically that question 79 was not applicable. Categories of only positive responses are as follows:

Tests set instructional goals and ensure coverage and pacing: (N=3)
"Teachers combined resources and ideas which helped me prepare meaningful teaching lessons."
"Knowing the scope and extent of info to be tested allowed me the opportunity to insure that these areas were taught or familiar by test time."

Identify weakness so as to refocus instruction (for the class as a whole): (N=3)
"I use the test results as one indicator of personal performance. Any time my average scores are not in stanines 7,8,9--I know I need to reevaluate and restructure my program."
"Testing allows me to determine which areas of my curriculum need improvement."

Identify strengths and weaknesses of individual students for extra help: (N=5)
"Test results give me a basic idea of what my students are weakest in. Then we can really work on these areas."

Improve achievement: (N=2)
"(The state test) improved writing skills for all my children."
"Improve reading stamina and learning more vocabulary words."

Motivate students: (N=2)
"Tests give children goals. They know what they have learned is acknowledged."

Other positive responses: (N=2)
e.g., "They gave me the idea of teaching beyond my grade level thus making it easier for my students in later years...."

Positive Responses: 54% of teachers (N=150) gave both positive and negative responses to questions 78 and 79 respectively. Categories of positive response are as follows:

Tests set instructional goals and ensure coverage and pacing: (N=18)
"They have given me a clearer notion of what it is I am supposed to teach."
"Standardized tests are used as a guide--like sets of objectives."
"Cover more material in less time."
"Standardized test gives me standards for teaching and a goal to work toward. It motivates me to use different teaching styles to effectively reach all of my children."
"I think the tests encourage teachers to insure learning."
"Standardized tests facilitate greater consistency in the curriculum both within classrooms at a single grade level as well as across grade levels school wide."
"Directs our attention to particular skills we use in our reading, writing, thinking, and discussion of a whole (integrated language arts program)."
"more intense teaching."

Ensure teaching of basic skills: (N=12)
"Standardized test gave a nice variety of the different basic skills that the children needed to master."
"The tests help us focus on certain basic skills that otherwise we may be some what relaxed (about)."
"Give me the incentive to drill on a daily basis--for a short, short time--fundamental facts."
"Puts emphasis on basic skills to that slower students master them."
"I probably put stronger emphasis on certain reading skills such as getting the main idea and reading for details in order to achieve a higher level of performance."
"Stress on standardized items and repeated repetition helps the slow learners in the group."
"It helps focus some teachers on basic skills who may otherwise not focus on anything at all."
"Because of these tests I am constantly reviewing and reinforcing skills taught all through the year."

Identify weakness so as to refocus instruction (for the class as a whole): (N=30)
"I use the results to help me identify the areas that my students are weak in, and focus on those areas first before implementing new concepts."
"The areas on which a large % score low (usage, etc.) I evaluated and restructured my instruction to improve student understanding and mastery."
"I paid closer attention to specific objectives and goals. I used it as a guide to personal improvement in teaching."
"They are one resource for feedback on what I need to emphasize more in my lesson planning."
"*Some* tests ask questions involving higher level thinking skills which require the teaching of those higher level thinking skills."
"I found out students needed more teaching of fractions and number sequences."
"Emphasizing skills where students were low--using different strategies to teach these skills. (District is getting better about offering training and encouraging the use of new strategies.)"
"It alerted me to focus on the skills the children were deficient in."

"After I have given pretests, I can analyze the results and concentrate on what areas are low."
"Identifying weakness, specifically in comprehension, has caused us to emphasize critical thinking skills and allowed greater flexibility in exploring new techniques."

Identify strengths and weakness of individual students for extra help: (N=39)
"Saw weaknesses in pretest and individualized instruction or used small group instruction to help master basics."
"When the test was used as a diagnostic tool to determine strengths and weaknesses."
"In some cases such tests have revealed weaknesses in time to remediate such weaknesses."
"It gave me a chance to know what to focus on if a child was weak in a particular area."
"Standardized tests do provide a starting point for emphasizing students' strengths and weaknesses. The tests also lead parents and teachers to discussions of individual students that can be helpful."
"In looking at test scores I am able to help student with the area that they are having the most trouble with."
"It helps to see what categories they are have mastered and those that are having trouble."
"It can highlight skills not mastered to be reviewed over the summer and during the upcoming year."
"The test showed (1) the areas where students need extra help (2) enabled the parents to share in upgrading their child's study habits."
"I emphasized word recognition skills after my ESL students did poorly."

Improve achievement: (N=6)
"Improves basic skills."
"(State test) scores went up dramatically."
"Reading skills and math skills were enhanced."

Writing test improves writing skills: (N=7)
"(State test) emphasis on writing results in constant writing efforts in all subject areas and development of a diversity of writing formats."
"It helped children improve and enjoy their writing."

Improve test-taking skills: (N=8)
"Test taking skills help those students who are afraid of test."
"Children get the experience in taking tests. They need to be test conscious. Society leans toward testing for many jobs."

Motivate students: (N=8)
"Children were anxious to learn to read in order to be able to take the test. They were eager to practice problem solving skills."
"Testing is a challenging follow up to instruction. Children love the competition."
"A form of goal, or incentive, to learn and do their best."

Improve student self-confidence: (N=5)
"It raised self-esteem tremendously in a couple of students who passed the tests much to their surprise."

"By taking the standardized tests the students felt like they were working toward a common goal. A hurdle to overcome. Getting high scores on the practice tests made them feel smarter."

Used in placement and grouping decisions: (N=2)
"Standardized tests are part of the prerequisite to get into the gifted class and advanced classes."
"Preparing my reading and math groups."

Provide norm-referenced information: (N=2)
"It helps to know where students rank among the nation."

Other positive responses: (N=13)
e.g., ""We team taught for (the state test) and found it good for students and teachers."
"It enables pupils to do critical thinking."
"They are simple straight to the point directions."
"Affirm current classroom results."

Negative Responses: **54%** of teachers (N=150) gave both positive and negative responses to questions 78 and 79 respectively. These are the same 150 teachers who gave positive answers above. Categories of negative response are as follows:

Too much teaching to test content and test format: (N=66)
"Teachers feel pressure to 'teach to test items' and avoid higher level thinking skills."
"I can't get to science and social studies, like I would like to. Instead, I'm preparing my class for a standardized test."
"Time I would like to have spent using the Whole Language approach was used up in practice tests and district practice materials. I felt pressure from my principal to complete all practice materials, because she feels that is the road to success."
"Critical thinking skills are basically non-existent in our children because of drill and practice for (Test 1), (Test 2), and (Test 3).
"You don't get to do as many special projects or have open discussions."
"I feel I could be more creative in my teaching if I didn't feel the necessity to be sure every child understood each basic skill."
"I wasn't able to do enrichment activities because they weren't part of the objectives. Students became burnt out on basic skills."
"Don't go through the curriculum in orderly fashion because trying to cover things on the test first. Cover some unimportant things because they are on the test."
"Teaching testing skills takes so much time that much other material cannot be taught."
"The students receive little hands-on learning in place of drill and specific skills teaching."
"Only test objectives are being taught. We need a well rounded curriculum."
"At times it does not allow me to be as creative as I wish. Especially when it comes to the writing. The tests expect so much structure."
"The timing of the tests in the year covering material not appropriate to be taught yet, yet needing to be mastered on the test."
"I spend less time teaching science and social studies because reading and math games are so important."

"Too much time is needed to emphasize test content, test taking skills, practice work-sheets."
"Limits creative teaching--You are always teaching (or so it seems) towards the test--Constant reminders on the P.A. system about test scores."
"Two weeks prior to testing an increase in practice on test format disrupts our usual unit plan, whole language, hands on approach to education."
"Constant drill, no higher cognitive skill learning for slow learners."
"Our lesson plans have to show what test objectives we're working on for the week. I'm forced to teach for the test."
"We spend too much time at our school on the basics and test taking techniques."
"Do not teach but basics on the test. Do not teach enough to excel in upper grades."
"I don't feel I should drill for the test format all year long."
"Had to move on too quickly in order to cover all material required on test."

Too much time, too much testing: (N=8)
"I have a poor attitude because here in (State) we test too much. There are State tests and national tests. We don't need so much nor so often. It makes me, and my students hostile toward testing."
"We spend too many days on standardized or district made test preparation materials."
"The time required to test detracted from time needed to teach--We used 10 days in fifth grade this year."

Test preparation boring to students: (N=4)
"Students and I become bored with practicing test taking skills."
"Kids start tuning out because they really dislike these tests (and that's with a positive approach from a teacher)."

Stressful for students and teachers: (N=28)
"Pressure to do well makes everyone uptight."
"Some of the children feared taking the tests, therefore they did not do as well as they were capable of doing."
"It tends to add a small amount of pressure on the students, especially when they know how to do a certain problem but can't remember."
"Too structured and takes 'fun' out of learning."
"The pressure on the students especially the younger ones of actually taking the test itself. Some were very nervous while others thought that the test was fun."
"Too much emphasis is put on student performance in negative ways such as newspaper reports. This causes student and teacher pressure and stress to overperform."
"Too much emphasis and pressure from the principal, district and media. In my opinion, the pressure encourages cheating from a lot of teachers."
"Mainly that the students and teachers get too stressed out before, during, and after the test."
"Students concentrate on 'what will be included on the test,' students are stressed out and get too nervous during the test."
"Students worrying about passing or doing well. Wasting time trying to comfort and telling them to do their best."
"These test scores are used to measure my effectiveness as a teacher in the classroom. The pressure from administration for high test scores and the media has a negative effect on teachers."

"Because the test results are 'flaunted' in the news etc., I become very resentful at having to give the tests. The pressure is bad on teachers and therefore passed on to students. It's as if school's out when tests are completed."
"The students are put under an unimaginable amount of pressure while taking the test and when results come back, depending on where they are, their level of esteem is considerably dropped.

Inaccurate measure of teachers, students, or curriculum: (N=24)
"These tests may test objectives not included in the curriculum."
"A student may have good knowledge of information but when it comes to testing, they can't demonstrate it as well as everyday classroom learning."
"I feel that when scores are published other factors should be included, i.e., economics, and social variables."
"(1)Comparison to others, negative effect on schools, districts, and students, (2) Year to year comparisons, i.e., growth, not including subjective factors like health, environment, etc.."
"The standardized test does not always reflect the capability of the student. Some students test well but will perform low and vice versa."
"These tests may test objectives not included in the curriculum."
"When a student scores high and is expected to function at that level, but the student can't actually perform at grade level."
"Tests inappropriate skills."
"Children that don't test well are penalized."
"A teacher in the previous grade teaches to the test and as a result when 3rd grade gets its children we see negative growth in many subjects each year. Although we know the explanation, this negative growth has shadowed our opinions of the jobs we've done...."

Negative influence on slow learners: (N=5)
"labels children."
"When students with excellent scores criticized students with lower scores, thus making them feel inferior."
"Standardized tests have had a negative influence on student learning when they are used to display their performance in a negative way rather than in a positive manner."

Culturally biased: (N=4)
"Cultural backgrounds being tested and compared--unfair to disadvantaged students."
"My bilingual students are sometimes not ready to take a standardized test but the district demands it. Students often guess and get frustrated."

Undermines teacher confidence: (N=4)
"Teacher morale! We have been harassed unmercifully this year, and it is wearing us out."
"When students still don't score as well even when you have been reteaching and going over skills consistently all year."

Other negative responses: (N=7)
"Publication in the newspaper of various schools' scores."
"The negative side of the issue is that the test results are used for all kinds of reasons besides helping the child."
"Some become frustrated during the actual test."

"Timed for mastery; some students require more time than most tests allocate. The reading passages are biased in reference to some students environment."

Only Negative: 24% (N=66) gave only negative answers, some said specifically that they could not think of positive answers to question 78. Categories of only negative responses are as follows:

Too much teaching to test content and test format: (N=28)
"Repetitive, boring drill. Use of district materials required too much time."
"Standardized test has had a negative influence on my teaching because some enrichment as well as practical skills must be set aside."
"I feel very sorry for the children in (State) whose school administrators and media cause them to miss out on so many areas of learning. Many teachers focus only on what tests cover, and I believe the children are suffering from it."
"I teach test taking skills because it's required."
"I was given an abundance of testing practice material each month and I was told I must do it."
"Too much time spent on test--not enough on thinking/problem solving."
"I had to spend a lot of time going over basic reading skills in all subject areas and basic skills in math and I felt I couldn't do a lot of fun math and science activities until the test was over."
"We are constantly reminded to practice, practice, for the test. The fun and excitement has been taken out of teaching."
"Many types of discovery learning cannot be assessed. Therefore not as much as should be is spent on this type of learning, e.g., science projects, group experiments."
"I always felt things were being rushed so much. Test objectives are becoming the subject matter and the activity of the day."

Too much time, too much testing: (N=6)
"I field-tested next year's (State test); students took one pretest for the district to prepare for (State test), the State test, and finally the (nationally normed test). Four weeks of testing is too much."
"Students became bored with the repetitive practice tests. Valuable class time was lost to practice test."

Stressful for students and teachers: (N=13)
"Highly stressful. I don't feel the results are accurate. Overwhelming demands and expectations on teachers."
"Everyone gets worked up about the test and misses the joy of learning. The test makers gear the test toward the W.A.S.P.s and exclude most other racial minorities."
"Too much pressure on the students and the teachers. There is not enough time to do extra supplemental activities and enrichment work."
"Students are very nervous and worried about the tests. They tend to shut down and not try their best."
"The teachers will cheat so that their students will score well, then go around bragging about it later. It creates a lot of animosity because everyone thinks that everyone else is cheating."
"Stress on teachers and students to be successful in a testing situation that shuts down the normal routine for children and teachers."
"Anxiety among the children; reduction in teacher choice; de-professionalization of teaching field."

Negative or unfair comparisons: (N=7)
"Our standardized test scores are typed up and passed around to every teacher to compare how they did to others. No mention goes to the teachers with low classes. It is much too competitive."
"When two rooms had low scores, all third grades suffered the indignity of having a weekly sit-in by a so-called '(Test)' helper who did nothing but sit."
"I do not feel standardized tests are used correctly--they should be used diagnostically not judgmentally. I should feel free to use them to help students, and not have them used against me. The public a press constantly tear schools apart because their test scores aren't as high as another. No one sees the fact that we don't all get students who are on the same level. I'm tired of being portrayed as an incompetent because of low scores."
"My children are intermediate bilingual and being compared with standard English speaking classes which have quite a head start. This is asking more than is reasonable for some of them."

Inaccurate measure of teachers, students, or curriculum: (N=6)
"Some of my students don't test well, although they are very creative and high achievers."
"Because of the extreme emphasis placed on these tests, I feel the results are not valid."
"Some people have test anxiety. The results are not related to their everyday skills. This can track a child and determine future placements."
"I feel that the tests we take are not testing what we should know at that grade level. Too much preparation time is expected."

Negative influence on slow learners: (N=3)
"Timed tests are very frustrating to a class of below average, at risk students. They need situations that are successful."
"The negative influence is reflected in the students' self concept. They compare their scores."

Undermines teacher confidence: (N=2)
"It is discouraging to me and I feel very constricted as a teacher. I'm the example: discouraged. I'm applying to another district."

Other negative responses: (N=1)